

Model identification using the Efficient Determination Criterion

Paulo Angelo Alves RESENDE^{a,*}, Chang Chung Yu DOREA^a

^a*Department of Mathematics, University of Brasília, Brasília 70910-900, Distrito Federal, Brazil*

Abstract

In the realm of the model selection context, Akaike's and Schwarz's information criteria, AIC and BIC, have been applied successfully for decades for model order identification. The Efficient Determination Criterion (EDC) is a generalization of these criteria, proposed originally to define a strongly consistent class of estimators for the dependency order of a multiple Markov chain. In this work, the EDC is generalized to partially nested models, which encompass many other order identification problems. Based on some assumptions, a class of strongly consistent estimators is established in this general environment. This framework is applied to BEKK multivariate GARCH models and, in particular, the strong consistency of the order estimator based on BIC is established for these models.

Keywords: EDC, BIC, AIC, Order estimation, BEKK-GARCH

2010 MSC: 62M05, 62F12, 60J10

1. Introduction

The order identification problem was initially dealt by using nested hypothesis tests in evaluating the order of multiple Markov chains [1, 2, 3, 4, 5], Autoregressive models [6, 7, 8, 9], among others. In the selection model context,
 5 Akaike [10] proposed the use of the information criterion AIC, aiming to avoid

*Corresponding author

Email addresses: pa@pauloangelo.com (Paulo Angelo Alves RESENDE),
chang@mat.unb.br (Chang Chung Yu DOREA)

empirical analysis on the estimation process. Subsequently, Schwarz [11] proposed the information criterion BIC. Since then, these criteria have been applied in such contexts as selecting models in Autoregressive (AR) and Autoregressive Moving Average (ARMA) process [12, 13, 14, 15, 16, 17], estimating dependency order in multiple Markov chains [18, 19, 20], detecting change-points in non-homogeneous Markov chains [21], estimating the length of the hidden state space of a hidden Markov model [22], estimating order in Autoregressive Conditional Heteroskedasticity process (ARCH) [23] and on estimating dependency order in specific situations [24].

15 Zhao et al. [25], on estimating the order of multiple Markov chains, introduced the Efficient Determination Criterion (EDC), which allows for adjustments on the penalty term used in the criteria AIC and BIC. Also, a class of strongly consistent estimators was established in the same work. Afterwards, Dorea [26] extended this class and proposed the asymptotic optimal order estimator, which had its better performance verified by the extensive use of numerical simulations [27].

In this work, the concept of “nested models” is generalized to class of partially nested models and the EDC criterion is extended to this new context. Some results regarding the consistency of EDC order estimators are established based essentially on assumptions about the likelihood function. This approach is applied to state the consistency of the BIC order estimator for BEKK multivariate GARCH models, which encompass the univariate version GARCH as particular case.

Section 2 provides the general results, that may be applied in a variety of models to establish the EDC order estimators. Section 3 presents the approach applied to BEKK multivariate GARCH models. The proofs of the stated results are in the appendices.

2. General framework

The essence of nested models have being used since the pioneer researches
 35 using hypothesis tests. However, practically all works focused on particular
 cases and the formal definition and treatment of the concept of nested models
 were unused. Nishii [28] firstly proposed a general estimator for the dimension of
 i.i.d. models. A relevant piece of Nishii's technique is adapted to our purposes.

For an arbitrary time discrete stochastic process $\mathbb{X} = \{X_t\}_{t \in \mathbb{N}}$, $E \subseteq \mathbb{R}^p$ the
 set of possible values of X_t and ν a fixed measure on E , we define a family of
 statistical models for \mathbb{X} as

$$M = \{f(x_1^n, \theta, n) : \theta \in \Theta, n \geq 1\}$$

where $f(x_1^n, \theta, n)$ represents the set of possible densities for x_1^n with respect to
 40 the product measure on E^n , which depends on the parameter $\theta \in \Theta \subseteq \mathbb{R}^d$, and
 $x_1^n = x_1 x_2 \dots x_n$ is a realization of \mathbb{X} . We may denote $f(x_1^n, \theta) = f(x_1^n, \theta, n)$ to
 simplify the notation.

Two statistical models

$$M_k = \{f(x_1^n, \theta, n) : \theta \in \Theta_k, n \geq 1\} \quad \text{and}$$

$$M_p = \{f(x_1^n, \theta, n) : \theta \in \Theta_p, n \geq 1\}$$

are nested, denoted by $M_k \subseteq M_p$, if $\Theta_k \subseteq \Theta_p$ and, for all $\theta \in \Theta_k$, $x_1^\infty \in E^\infty$,
 exists $c \in (0, \infty)$ such as

$$\lim_{n \rightarrow \infty} \frac{f_k(x_1^n, \theta)}{f_p(x_1^n, \theta)} = c.$$

For $q \in \mathbb{N}$, $p = (p_1, \dots, p_q) \in \mathbb{N}^q$ and $k = (k_1, \dots, k_q) \in \mathbb{N}^q$, we define the usual
 order relation $p \geq k$ iff $p_i \geq k_i$ for $i = 1 \dots q$, which makes (\mathbb{N}, \geq) a partially
 ordered set. For $p \not\geq k$ we mean that $p < k$ or p and k are not related. The set
 $\mathbb{M} = \{M_k\}_{k \in \mathbb{N}^q}$ is a class of partially nested models if

$$M_k \subseteq M_p \Leftrightarrow k \leq p.$$

We say that an element $m_r \in \bigcup_{k \in \mathbb{N}^q} M_k$ has order $r \in \mathbb{N}^q$ if $m_r \in M_r$ and $m_r \in$
 M_k implies that $M_r \subseteq M_k$. In this context, for a sample x_1^n and $\hat{\theta}_k$ the maximum

45 likelihood estimator (MLE) of θ supposing the order k , $\{L_{n,k}(x_1^n, \theta)\}_{(n,k) \in \mathbb{N} \times \mathbb{N}^q}$
is a class of functions $L_{n,k} : E^n \times \Theta_k \rightarrow \mathbb{R}$ that satisfies

$$L_{n,k}(x_1^n, \hat{\theta}_k) = \sup_{\theta \in \Theta_k} \{L_{n,k}(x_1^n, \theta)\} \quad (1)$$

and for $\theta \in \Theta_k$ and $p \geq k$,

$$L_{n,p}(x_1^n, \theta) \geq L_{n,k}(x_1^n, \theta) \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{L_{n,p}(x_1^n, \theta)}{L_{n,k}(x_1^n, \theta)} < \infty. \quad (2)$$

To simplify notation, we shall denote $L_{n,k}(\theta) = L_{n,k}(x_1^n, \theta)$. In most situation,
the $L_{n,k}$ functions are merely the likelihood for each n and k . Now, we define
50 the EDC estimator for a class of partially nested models.

Definition 1. Let \mathbb{M} be a class of partially nested models, $m_r \in \bigcup_{k \in \mathbb{N}^q} M_k$ of
order r and $K \geq r$. The EDC estimator is defined by

$$\hat{r} = \operatorname{argmin}_{k \leq K} \{\operatorname{EDC}(k)\} \quad (3)$$

for

$$\operatorname{EDC}(k) = -\log L_{n,k}(\hat{\theta}_k) + c_n \gamma(k),$$

c_n a sequence of positive numbers and $\gamma(k) = \dim(\Theta_k)$.

We need the following assumptions to conclude consistency for the EDC es-
55 timator based on the asymptotic behaviour of the c_n sequence. In what follows,
 r is the order of \mathbb{X} , θ_r is the true parameter, i.e. the one that gives the density
for the process \mathbb{X} and $\hat{\theta}_k$ is the MLE of θ_r supposing the order k .

Assumption A1. For all $k \geq r$, θ_r is an interior point of Θ_k and

$$\hat{\theta}_k \xrightarrow{a.s.} \theta_r.$$

Assumption A2. For all $k, n \in \mathbb{N}$, $\log L_{n,k}(x_1^n, \theta)$ and its derivatives
 $D_\theta^1(\log L_{n,k}(x_1^n, \theta))$, $D_\theta^2(\log L_{n,k}(x_1^n, \theta))$ and $D_\theta^3(\log L_{n,k}(x_1^n, \theta))$ are measurable
60 with respect to x_1^n and continuous with respect to θ .

Assumption A3. For $k \geq r$, there exists $c < \infty$ and a symmetric and positive definite matrix A_2 , such as, for all $\dot{\theta} = (1-s)\hat{\theta}_k + s\theta_r$, $s \in (0,1)$, $(i,j,l) \in \{1, \dots, \gamma(k)\}^3$,

$$\lim_{n \rightarrow \infty} \frac{\left(D_{\dot{\theta}}^3(\log L_{n,k}(\dot{\theta})) \right)_{i,j,l}}{n} < c \quad a.s. \quad \text{and}$$

$$\lim_{n \rightarrow \infty} \frac{D_{\dot{\theta}}^2(\log L_{n,k}(\dot{\theta}))}{n} = A_2 \quad a.s.$$

Assumption A4. If $k \geq r$,

$$\limsup_{n \rightarrow \infty} \left\| \frac{D_{\theta}^1 \log L_{n,k}(\theta_r)}{\sqrt{2n \log \log n}} \right\| < \infty \quad a.s. \quad (4)$$

Assumption A5. If $k \not\geq r$,

$$0 < \lim_{n \rightarrow \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,p}(\hat{\theta}_k)}{n} \quad a.s.$$

We note that Basawa & Heyde [29] propose basically the use of assumptions A1-A3 and an analogous of A4 to conclude the asymptotically normality of the parameter estimator $\hat{\theta}_k$. The approach used here is quite similar to their. The next result establishes the class of consistent EDC order estimators based on the assumptions and on the asymptotic behaviour of the sequence c_n . The proof is in the Appendix A. The Corollary concludes the strong consistency for the BIC order estimator defined in this general context.

Theorem 1. Let \mathbb{X} be a discrete time stochastic process taking values in \mathbb{R}^m , \mathbb{M} its respective class of partially nested models, $m_r \in \bigcup_{k \in \mathbb{N}^q} M_k$ of order r , \hat{r} as defined in (3), and the assumptions A1-A5 are satisfied. Then $\hat{r} \xrightarrow{a.s.} r$ if

$$\liminf_{n \rightarrow \infty} \frac{c_n}{\log \log n} = \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{c_n}{n} = 0.$$

Corollary 1. Supposing the same hypothesis of Theorem 1, the BIC order estimator bellow is strongly consistent.

$$\hat{r}_{bic} = \operatorname{argmin}_{k \leq K} \left\{ -\log L_{n,k}(\hat{\theta}_k) + \frac{\log n}{2} \gamma(k) \right\}$$

for a known $K \geq r$.

70 The proof consists in the determination of the asymptotic behaviour of the differences bellow for arbitrary $p \not\geq r$ and $k > r$.

$$\frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,p}(\hat{\theta}_p)}{n} \quad (5)$$

$$\frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,r}(\hat{\theta}_r)}{\log \log n} \quad (6)$$

Assumption A5 is precisely (5). For (6) we use assumptions A1 and A2 to enable the use of Taylor series and state

$$\limsup_{n \rightarrow \infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,k}(\theta_r)}{\log \log n} \leq \limsup_{n \rightarrow \infty} \frac{\sqrt{n}(\theta_r - \hat{\theta}_k)}{\sqrt{2 \log \log n}} \frac{-D_{\theta}^2 \left(\log L_{n,k}(\hat{\theta}_k) \right)}{n} \frac{\left(\sqrt{n}(\theta_r - \hat{\theta}_k) \right)'}{\sqrt{2 \log \log n}} + o(\log \log n)$$

75 and

$$\frac{1}{\sqrt{n \log \log n}} D_{\theta}^1 \log L_{n,k}(\theta_r) \left(\frac{-D_{\theta}^2 \left(\log L_{n,k}(\hat{\theta}_k) \right)}{n} \right)^{-1} = \frac{\sqrt{n}}{\sqrt{\log \log n}} (\hat{\theta}_k - \theta_r).$$

Using A3 and A4 we define a upper bound for (6).

Defining the asymptotic behaviour of $D_{\theta}^1 \log L_{n,k}(\theta_r)$ is generally easier when compared to the effort in manipulating directly the equation (6) to state its asymptotic behaviour. Besides that, assumptions A1-A3 are commonly used to
80 establish the asymptotic normality for the MLE and usually are available in the literature.

3. BEKK-GARCH order estimation

Engle [30] originally proposed the use of ARCH models on modelling time series in economy. His work has been hugely influential in the area and motivated
85 many generalizations and/or adaptations such as GARCH [31], NGARCH [32], EGARCH [33] and the multivariate generalizations BEKK-GARCH [34], VEC-GARCH [35], CCC-GARCH [36], Factor-GARCH [37], among others.

The multivariate models have special applications in portfolio selection and asset pricing. In this family, the BEKK-GARCH models has particular relevance
90 due to its generality and the amount of research available in the literature.

Among the mentioned, only the VEC-GARCH is more general than the BEKK-GARCH model. However, the VEC-GARCH cases that can not be represented in the BEKK-GARCH parametrization are somewhat degenerated [38, 39].

Boussama [40] immersed the BEKK-GARCH models into the framework of
95 general state space Markov chains and used algebraic topology to conclude the geometric ergodicity of such models under regularity conditions. This work is also published in [38] with minor changes.

Comte & Lieberman [41] used Boussama's results to prove the consistency conditions proposed by Jeantheau [42] and the conditions proposed by Basawa
100 & Heyde [29] and conclude the strong consistency and asymptotic normality for the MLE of the parameter θ_r .

As with the particular case of ARCH models, until now, there is no results regarding consistency of order estimators for BEKK-GARCH models. However, the AIC and BIC information criteria have been used without further formal-
105 ization. In what follows, we present some preliminary results, which are used to prove the assumptions A1-A5 and conclude the consistency of the EDC order estimator \hat{r} for such models, which encompass the consistency of the BIC order estimator as a particular case.

For $k = (k_1, k_2) \in \mathbb{N}^2$, a random sequence $\mathbb{X} = \{X_t\}_{t \in \mathbb{N}}$ taking values in \mathbb{R}^m
110 is a BEKK-GARCH(k) model if for all $t \in \mathbb{N}$,

$$X_t = (H_t)^{\frac{1}{2}} \varepsilon_t, \quad (7)$$

where, for $m \times m$ matrices C , $\{A_{ls}\}$ and $\{B_{ls}\}$, C positive definite and $N \in \mathbb{N}$,

$$H_t = C + \sum_{l=1}^{k_2} \left(\sum_{s=1}^N A_{ls} X_{t-l} X'_{t-l} A'_{ls} \right) + \sum_{l=1}^{k_1} \left(\sum_{s=1}^N B_{ls} H_{t-l} B'_{ls} \right),$$

$\{\varepsilon_t\}_{t \in \mathbb{N}} \sim \mathcal{N}(0, I_m)$, and I_m is the $m \times m$ identity matrix. The process \mathbb{X} can be represented as a Markov chain $\mathbb{Y} = \{Y_t\}_{t \in \mathbb{N}}$ defined by

$$Y_t = (\text{vech}(H_{t+1})', \text{vech}(H_t)', \dots, \text{vech}(H_{t-k_1+2})', X'_t, X'_{t-1}, \dots, X'_{t-k_2+1})',$$

where vech is the operator that stacks the lower triangular portion of a matrix. Boussama proved that \mathbb{Y} is a positive Harris and geometric ergodic Markov

chain if

$$\rho \left(\sum_{l=1}^{k_2} \tilde{A}_l + \sum_{l=1}^{k_1} \tilde{B}_l \right) < 1, \quad (8)$$

where ρ is the spectral radius,

$$\tilde{A}_l = D_m^+ \sum_{s=1}^N (A_{ls} \otimes A_{ls}) D_m \quad \text{and} \quad \tilde{B}_l = D_m^+ \sum_{s=1}^N (B_{ls} \otimes B_{ls}) D_m,$$

vec is the operator that stacks the columns of a matrix, \otimes is the Kronecker product, D_m and D_m^+ are defined by the operators that satisfy

$$vec(A) = D_m vech(A) \quad \text{and}$$

$$vech(A) = D_m^+ vec(A).$$

For $k = (k_1, k_2)$ and a fixed $N \geq 1$, the BEKK-GARCH(k) model can be immersed in a class of partially ordered models considering, for $p > k$ and $\theta_k \in \Theta_k$, $\Theta_k \subsetneq \Theta_p$, the matrices $\{A_{ls}(\theta_k)\}$, $\{B_{l's}(\theta_k)\}$ and $C(\theta_k)$ for $l = 1 \cdots k_2$, $l' = 1 \cdots k_1$ and $s = 1 \cdots N$. Denoting $\bar{k} = \max(k_1, k_2)$, we have

$$f(x_1^n, \theta_k) = C_1(x_1^{\bar{k}}) \prod_{t=1+\bar{k}}^n \frac{1}{\sqrt{(2\pi)^{m/2} \det(H_{t,\theta_k})}} \exp \left(-\frac{1}{2} x_t' H_{t,\theta_k}^{-1} x_t \right)$$

for

$$H_{t,\theta_k} = C + \sum_{l=1}^{k_2} A_l X_{t-l} X_{t-s}' A_l' + \sum_{l=1}^{k_1} B_l H_{t-l,\theta_k} B_l'.$$

Also, the following definition for the functions $\log L_{n,k}$ satisfies (1) and (2).

$$\log L_{n,k}(\theta_k) = \sum_{1+\bar{k}}^n l_t(\theta_k)$$

for

$$l_t(\theta_k) = -\frac{1}{2} X_t' H_{t,\theta_k}^{-1} X_t - \frac{1}{2} \log \det(H_{t,\theta_k}). \quad (9)$$

115 The nesting relation $M_k \subset M_p$ can be observed taking A_{ls} and $B_{l's}$ as null matrices for $l > k_2$ and $l' > k_1$.

In particular, if $N = 1$ and θ_k is the columns of the matrices $\{A_l\}$, $\{B_{l'}\}$ and C , we may construct $\Theta_k \subseteq \mathbb{R}^{m^2(2\bar{k}+1)}$,

$$\Theta_k = \Omega_0 \times \Omega_1 \times \cdots \times \Omega_{2\bar{k}},$$

for $\Omega_i = \{0\}^{m^2}$ if $i/2 > k_2$ and i is odd or $i/2 > k_1$ and i is even, for the remaining cases, $\Omega_i \subseteq \mathbb{R}^{m^2}$ has non-empty interior. Assuming $A_i = 0$ if $i > k_2$, $B_i = 0$ if $i > k_2$,

$$\theta_k = (\text{vec}(C), \text{vec}(A_1), \text{vec}(B_1), \dots, \text{vec}(A_{\bar{k}}), \text{vec}(B_{\bar{k}})) \in \Theta_k.$$

In this case,

$$\gamma(k) = m^2(1 + k_1 + k_2).$$

For the order, r , of a BEKK-GARCH process \mathbb{X} , we consider the lowest k such as \mathbb{X} can be represented by (7). Assuming the following conditions (B1-B5), we establish the Theorem 2 that concludes assumptions A1-A5 and states
120 the class of strong consistent EDC order estimators. For $k \geq r$,

B1. Θ_k is compact and θ_r is an interior point of Θ_k .

B2. There exists a $c > 0$ such as $\inf_{\theta \in \Theta_k} \det C(\theta) > c$.

B3. The model is identifiable, i.e. $H_t(\theta) = H_t(\theta')$ a.s. if and only if $\theta = \theta'$.

B4. $C(\theta)$, $\tilde{A}_l(\theta)$ and $\tilde{B}_l(\theta)$ and their derivatives, with respect to θ , until order
125 3 are continuous.

B5. X_t admits bounded moments of order 16.

Comte & Lieberman use B1-B4 and finite moments of order 8 in B5 to conclude the asymptotic normality of $\hat{\theta}_k$. We need the finiteness for moments of order 16 to conclude assumption A4.

Theorem 2. *Let \mathbb{X} be a BEKK-GARCH(r) of order r , satisfying (8) and conditions B1-B5. Then the EDC order estimator defined at (3) is strongly consistent if*

$$\liminf_{n \rightarrow \infty} \frac{c_n}{\log \log n} = \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{c_n}{n} = 0.$$

130 Boussama [40] concluded the geometric ergodicity for the associated Markov chain of \mathbb{X} and, in particular, enabled the use of the Strong Law of Large Numbers (SLLN), which can be found in Meyn & Tweedie [43]. However, the geometric ergodicity is not sufficient to conclude the Law of Iterated Logarithm

(LIL), needed to prove assumption A4. To overcome this, we use the LIL below, stated for square integrable Martingales, which can be found in [44]. Also, some auxiliary results stated by Comte & Lieberman [41] are used to conclude assumptions A1-A5.

Theorem 3 (Hall & Heyde (1980)). *Let $\{S_n, \mathcal{F}_{t-1}\}$ be a martingale, $S_n = \sum_{t=1}^n U_t$, $E(S_n) = 0$, $E(S_n^2) < \infty$, $\{Z_t\}_{t \in \mathbb{N}}$ and $\{W_n\}_{n \in \mathbb{N}}$ non-negative random variables such as Z_t and W_t are \mathcal{F}_{t-1} measurable. If*

$$\lim_{n \rightarrow \infty} \frac{\sum_{t=1}^n U_t \mathbb{I}(|U_t| > Z_t) - E[U_t \mathbb{I}(|U_t| > Z_t) | \mathcal{F}_{t-1}]}{\sqrt{2W_n^2 \log \log W_n^2}} = 0 \quad a.s., \quad (\text{L1})$$

$$\lim_{n \rightarrow \infty} \frac{\sum_{t=1}^n E[U_t^2 \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}] - E[U_t \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}]^2}{W_n^2} = 1 \quad a.s., \quad (\text{L2})$$

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n \frac{E[U_t^4 \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}]}{W_t^4} < \infty \quad a.s., \quad (\text{L3})$$

$$\lim_{n \rightarrow \infty} \frac{W_n}{W_{n+1}} = 1 \quad a.s. \quad \text{and} \quad \lim_{n \rightarrow \infty} W_n = \infty \quad a.s. \quad (\text{L4})$$

Then

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2W_n^2 \log \log W_n^2}} = 1 \quad a.s.$$

and

$$\liminf_{n \rightarrow \infty} \frac{S_n}{\sqrt{2W_n^2 \log \log W_n^2}} = -1 \quad a.s.$$

4. Conclusion

The Efficient Determination Criterion (EDC) raises as a promising approach in the context of partially nested models. Mainly because the assumptions A1-A5 simplify the establishment of strongly consistent order estimators in a variety of models. Some of these assumptions, for each case, can be found in the literature on defining the asymptotic normality for the respective MLE.

150 Hafner & Preminger [45] state some results for VEC-GARCH(1,1) models.
 If it is possible to generalize these results for arbitrary $k \in \mathbb{N}^2$, the EDC order
 estimator can be easily defined for VEC-GARCH models.

As future works, we suggest to weaken the hypothesis B5 and state the
 consistency of the EDC estimator for $c_n = O(\log \log n)$.

155 Acknowledgement

The authors gratefully acknowledge the attention of Dr. Farid Boussama in
 sending his thesis and indirectly encouraging the establishment of the EDC for
 BEKK-GARCH models.

Appendix A. Proof of Theorem 1

160 The following Lemma is an adaptation of results that can be found at [26].

Lemma 1. *Let \mathbb{X} be a discrete time stochastic process with values in \mathbb{R}^m , \mathbb{M}
 its respective class of partially nested models, $m_r \in \bigcup_{k=0}^{\infty} M_k$ of order r and \hat{r}
 as defined in (3). Then \hat{r} is strongly consistent ($\hat{r} \xrightarrow[a.s.]{} r$) if, for $k \not\geq r$, exists
 $c_1 \in (0, \infty)$ such as*

$$\lim_{n \rightarrow \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,k}(\hat{\theta}_k)}{n} \geq c_1 \text{ a.s.}, \quad (\text{A.1})$$

165 for $k > r$, exists $c_2 \in (0, \infty)$ such as

$$\limsup_{n \rightarrow \infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,r}(\hat{\theta}_r)}{\log \log n} \leq c_2(\gamma(k) - \gamma(r)) \text{ a.s.} \quad (\text{A.2})$$

and c_n satisfies

$$\lim_{n \rightarrow \infty} \frac{c_n}{n} = 0 \quad \text{and} \quad \liminf_{n \rightarrow \infty} \frac{c_n}{\log \log n} \geq c_2. \quad (\text{A.3})$$

Proof. We have that

$$\begin{aligned} & \left(-\log L_{n,p}(\hat{\theta}_p) + \gamma(p)c_n \right) - \left(-\log L_{n,l}(\hat{\theta}_l) + \gamma(l)c_n \right) \\ &= \left(\log L_{n,l}(\hat{\theta}_l) - \log L_{n,p}(\hat{\theta}_p) \right) - c_n (\gamma(l) - \gamma(p)). \end{aligned} \quad (\text{A.4})$$

Taking $p = r$ and $l = k$ in (A.4) and using (A.2) we get

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{(-\log L_{n,k}(\hat{\theta}_r) + \gamma(r)c_n) - (-\log L_{n,k}(\hat{\theta}_k) + \gamma(k)c_n)}{\log \log n} \\ \leq c_2(\gamma(k) - \gamma(r)) - \liminf_{n \rightarrow \infty} \left(\frac{c_n}{\log \log n} \right) (\gamma(k) - \gamma(r)) \quad a.s. \\ \leq c_2(\gamma(k) - \gamma(r)) - c_2(\gamma(k) - \gamma(r)) \quad a.s. \\ = 0. \end{aligned}$$

In the same manner, but taking $l = r$ and $p = k < r$ in (A.4), and using (A.1) and (A.3), we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{(-\log L_{n,k}(\hat{\theta}_k) + \gamma(k)c_n) - (-\log L_{n,k}(\hat{\theta}_r) + \gamma(r)c_n)}{n} \\ \geq c_1 - \limsup_{n \rightarrow \infty} \frac{c_n}{n} (\gamma(r) - \gamma(k)) \quad a.s. \\ > 0 \quad a.s. \end{aligned}$$

Then, using its definition, we conclude that $\hat{r} \xrightarrow{a.s.} r$. \square

Lemma 2. Let \mathbb{X} be a discrete time stochastic process with values in \mathbb{R}^m , r its order, \mathbb{M} its respective class of partially nested models, the $\log L_{n,k}$ functions as defined above and $\hat{\theta}_k \in \Theta_k$ the MLE of the true parameter $\theta_r \in \Theta_k$. If assumptions A1-A5 are true, then

$$\limsup_{n \rightarrow \infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,r}(\hat{\theta}_r)}{\log \log n} \leq \frac{2c^2}{\lambda_{\gamma(k)}} \quad a.s.,$$

where $\lambda_{\gamma(k)}$ is the lowest eigenvalue of A_2 and c is an upper bound for (4).

Proof. Using A1-A2, for large enough n , we can take the Taylor expansion of $\log L_{n,k}(\theta_r)$ at $\hat{\theta}_k$, which gives

$$\begin{aligned} \log L_{n,k}(\theta_r) &= \log L_{n,k}(\hat{\theta}_k) + (\theta_r - \hat{\theta}_k) D_{\theta}^1 \left(\log L_{n,k}(\hat{\theta}_k) \right) \\ &\quad + \frac{1}{2} (\theta_r - \hat{\theta}_k) D_{\theta}^2 \left(\log L_{n,k}(\hat{\theta}_k) \right) (\theta_r - \hat{\theta}_k)^T + r_n (\theta_r - \hat{\theta}_k) \quad (A.5) \end{aligned}$$

where, for $\theta_r = (\alpha_1, \dots, \alpha_{\gamma(k)})$, $\hat{\theta}_k = (\hat{\alpha}_1, \dots, \hat{\alpha}_{\gamma(k)})$ and $\dot{\theta} = (1-s)\theta_k + s\theta_r$, $s \in (0, 1)$,

$$r_n (\theta_r - \hat{\theta}_k) = \frac{1}{3!} \sum_{i,j,l} \left(D_{\theta}^3 \log L_{n,k}(\dot{\theta}) \right)_{i,j,l} (\alpha_i - \hat{\alpha}_i) (\alpha_j - \hat{\alpha}_j) (\alpha_l - \hat{\alpha}_l).$$

By definition, $\hat{\theta}_k$ maximizes $L_{n,k}$, which gives $D_{\theta}^1 \left(\log L_{n,k}(\hat{\theta}_k) \right) = 0$. Organizing (A.5) and dividing by $\log \log n$, we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,k}(\theta_r)}{\log \log n} &\leq \limsup_{n \rightarrow \infty} \frac{\sqrt{n}(\theta_r - \hat{\theta}_k) - D_{\theta}^2 \left(\log L_{n,k}(\hat{\theta}_k) \right)}{\sqrt{2 \log \log n}} \frac{\left(\sqrt{n}(\theta_r - \hat{\theta}_k) \right)^T}{\sqrt{2 \log \log n}} \\ &\quad + \limsup_{n \rightarrow \infty} \frac{|r_n(\theta_r - \hat{\theta}_k)|}{\log \log n}. \end{aligned} \quad (\text{A.6})$$

Now, taking the Taylor expansion of $D_{\theta}^1 \log L_{n,k}(\hat{\theta}_k)$ at θ_r ,

$$\begin{aligned} (0, \dots, 0) &= D_{\theta}^1 \log L_{n,k}(\hat{\theta}_k) \\ &= D_{\theta}^1 \log L_{n,k}(\theta_r) + (\hat{\theta}_k - \theta_r) D_{\theta}^2 \log L_{n,k}(\dot{\theta}), \end{aligned}$$

180 where $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_k$ and $s \in (0, 1)$. Organizing, we have

$$\begin{aligned} \frac{1}{\sqrt{n \log \log n}} D_{\theta}^1 \log L_{n,k}(\theta_r) &= \frac{\sqrt{n}}{n \sqrt{\log \log n}} \left\{ (\hat{\theta}_k - \theta_r) D_{\theta}^2 \log L_{n,k}(\dot{\theta}) \right\} \\ &= -\frac{\sqrt{n}}{\sqrt{\log \log n}} (\hat{\theta}_k - \theta_r) \left[\frac{D_{\theta}^2 \log L_{n,k}(\dot{\theta})}{n} \right]. \end{aligned}$$

Using that A_2 is positive definite we conclude that it is invertible and, for large enough n ,

$$A_n := - \left[\frac{D_{\theta}^2 \log L_{n,k}(\dot{\theta})}{n} \right]$$

has inverse A_n^{-1} , then

$$\frac{1}{\sqrt{n \log \log n}} D_{\theta}^1 \log L_{n,k}(\theta_r) A_n^{-1} = \frac{\sqrt{n}}{\sqrt{\log \log n}} (\hat{\theta}_k - \theta_r). \quad (\text{A.7})$$

Using A4 and (A.7), considering $P_i : \mathbb{R}^{\gamma(k)} \rightarrow \mathbb{R}$ as the projection of coordinate i , we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left| \frac{\sqrt{n}(\hat{\alpha}_i - \alpha_i)}{\sqrt{2 \log \log n}} \right| &= \limsup_{n \rightarrow \infty} \left| P_i \left(\frac{\sqrt{n}}{\sqrt{2 \log \log n}} (\hat{\theta}_k - \theta_r) \right) \right| \\ &= \limsup_{n \rightarrow \infty} \left| P_i \left(\frac{1}{\sqrt{2n \log \log n}} D_{\theta}^1 \log L_{n,k}(\theta_r) A_n^{-1} \right) \right| \\ &= \limsup_{n \rightarrow \infty} \left| P_i \left(\frac{1}{\sqrt{2n \log \log n}} D_{\theta}^1 \log L_{n,k}(\theta_r) A_2^{-1} \right) \right| \\ &< \infty, \end{aligned}$$

which gives, using A1 and A2,

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \frac{|r_n(\theta_r - \hat{\theta}_k)|}{\log \log n} &\leq \limsup_{n \rightarrow \infty} \frac{1}{3!} \sum_{i,j,l} \left| \frac{(D_\theta^3 \log L_{n,k}(\hat{\theta}))_{i,j,l}}{n} \right| \left| \frac{\sqrt{n}(\alpha_i - \hat{\alpha}_i)}{\sqrt{\log \log n}} \right| \left| \frac{\sqrt{n}(\alpha_j - \hat{\alpha}_j)}{\sqrt{\log \log n}} \right| |\alpha_l - \hat{\alpha}_l| \\
&\leq c \sum_l \limsup_{n \rightarrow \infty} |\alpha_l - \hat{\alpha}_l| = 0 \quad a.s.
\end{aligned} \tag{A.8}$$

185 Using A3, (A.6), (A.7), (A.8) and that $A_n \xrightarrow{a.s.} A_2$, we have

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,k}(\theta_r)}{\log \log n} &\leq \limsup_{n \rightarrow \infty} \frac{(D_\theta^1 \log L_{n,k}(\theta_r)) A_n^{-1} - D_\theta^2 (\log L_{n,k}(\hat{\theta}_k))}{\sqrt{2n \log \log n}} \left(\frac{(D_\theta^1 \log L_{n,k}(\theta_r)) A_n^{-1}}{\sqrt{2n \log \log n}} \right)^T \\
&= \limsup_{n \rightarrow \infty} \frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n \log \log n}} A_2^{-1} A_2 \left(\frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n \log \log n}} A_2^{-1} \right)^T \\
&= \limsup_{n \rightarrow \infty} \frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n \log \log n}} A_2^{-1} \left(\frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n \log \log n}} \right)^T \\
&\leq \frac{1}{\lambda_{\gamma(k)}} \limsup_{n \rightarrow \infty} \left\| \frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n \log \log n}} \right\|^2 \\
&\leq \frac{c_5^2}{\lambda_{\gamma(k)}} \quad a.s.
\end{aligned} \tag{A.9}$$

We used that A_2 is symmetric positive definite so is its inverse. Applying (A.9) twice, we conclude the proof. \square

Using Assumption A5 and Lemma 2 we have (A.1) and (A.2). The Theorem follows from Lemma 1.

190 Appendix B. Proof of Theorem 2

Lemma 3 (Comte & Lieberman (2003)). *Let $\mathbb{X} = \{X_t\}_{t \in \mathbb{N}}$ be a BEKK-GARCH(k), $\theta_r = (\alpha_1, \dots, \alpha_{\gamma(k)})$ its true parameter, $\hat{\theta}_k = (\hat{\alpha}_1, \dots, \hat{\alpha}_{\gamma(k)})$ the MLE of θ_r . If conditions (8) and B1-B5 are true, then*

$$(i) \quad \lim_{n \rightarrow \infty} -\frac{D_\theta^2 \log L_{n,k}(\theta_r)}{n} = A_2 \quad a.s.,$$

where

$$A_2 = -E \left(\frac{\partial^2 l_t(\theta_r)}{\partial \theta \partial \theta'} \right). \tag{B.1}$$

(ii) A_2 is positive definite.

(iii) For all $i, j, l \in \{1, \dots, \gamma(k)\}$,

$$E \left(\sup_{\|\theta - \theta_r\| \leq \delta} \left| \frac{\partial^3 l_t(\theta)}{\partial \alpha_i \alpha_j \alpha_l} \right| \right) < c(\delta).$$

(iv) For all $i \in \{1, \dots, \gamma(k)\}$, $\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}$ is a square-integrable Martingale.

(v) The MLE $\hat{\theta}_k$ is strongly consistent.

(vi) Exists $c \in (0, \infty)$, which does not depend on t or θ , such as

$$\|H_t^{-1}\| \leq c.$$

(vii)

$$E(|\log(\det(H_t(\theta_r)))|) < \infty.$$

The following Lemma adapts some results from Comte & Lieberman to our purposes.

Lemma 4. Let $\mathbb{X} = \{X_t\}_{t \in \mathbb{N}}$ be a BEKK-GARCH(k), $\theta_r = (\alpha_1, \dots, \alpha_{\gamma(k)})$ its true parameter, $\log L_{n,k}$ as defined early, $\hat{\theta}_k \in \Theta_k$ the MLE of θ_r , $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_k$ and $s \in [0, 1]$ and $B_\delta(\theta_r) \subset \Theta_k$ a neighborhood of θ_r . If conditions (8) and B1-B5 are true, then

(i) Exists $c \in (0, \infty)$, such as, for all $i, j, l \in \{1, \dots, \gamma(k)\}$,

$$\limsup_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in B_\delta(\theta_r)} \frac{\partial^3 l_t(\theta)}{\partial \alpha_i \alpha_j \alpha_l} \right\| \leq c.$$

(ii)

$$\lim_{n \rightarrow \infty} -\frac{D_\theta^2 \log L_{n,k}(\dot{\theta})}{n} = A_2 \quad a.s.$$

for A_2 as defined in (B.1).

(iii)

$$E(|\log[\det(H_t(\theta_r))] + X_t' H_t^{-1} X_t|) < \infty.$$

205 *Proof.* (i) Using item (iii) of Lemma 3 and the Boussama's results, we just apply the SLLN that can be found at [43].

(ii) Analogous to the technique used in Lemma 5 of [46], using that $D_\theta^2 l_t(\theta)$ and $D_\theta^3 l_t(\theta)$ are continuous with respect to θ , that $\hat{\theta}_k$ is strongly consistent (Lemma 3) and the mean value Theorem, we have

$$\left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 l_t(\dot{\theta})}{\partial \alpha_i \alpha_j} - \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 l_t(\theta_r)}{\partial \alpha_i \alpha_j} \right\| \leq \sup_{\theta \in B_\delta(\theta_r)} \left\{ \left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial}{\partial \theta'} \left(\frac{\partial^2 l_t(\theta)}{\partial \alpha_i \alpha_j} \right) \right\| \cdot \|\dot{\theta} - \theta_r\| \right\}.$$

Using item (i) and the strong consistency of $\hat{\theta}$ we conclude the result.

(iii)

$$\begin{aligned} E(|\log[\det(H_t(\theta_r))] + X_t' H_t^{-1} X_t|) &\leq E(|\log[\det(H_t(\theta_r))]|) + E(|X_t' H_t^{-1} X_t|) \\ &\leq E(|\log[\det(H_t(\theta_r))]|) + E(\|X_t\|^2) E(\|H_t^{-1}\|) \end{aligned}$$

which is bounded by items (vi) and (vii) of Lemma 3 and by B3.

□

Lemma 5. Let $\mathbb{X} = \{X_t\}_{t \in \mathbb{N}}$ be a BEKK-GARCH(k), $\theta_r = (\alpha_1, \dots, \alpha_{\gamma(k)})$ its true parameter, If conditions (8) and B1-B5 are true, thus, for all $i \in \{1, \dots, \gamma(k)\}$,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}}{\sqrt{2n \log \log n}} &= E \left(\frac{\partial l_1(\theta_r)^2}{\partial \alpha_i} \right)^{1/2} \quad a.s., \\ \liminf_{n \rightarrow \infty} \frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}}{\sqrt{2n \log \log n}} &= -E \left(\frac{\partial l_1(\theta_r)^2}{\partial \alpha_i} \right)^{1/2} \quad a.s. \quad \text{and} \\ \limsup_{n \rightarrow \infty} \frac{\|D_\theta^1 \log L_{n,k}(\theta_r)\|}{\sqrt{2 \log \log n}} &\leq c \quad a.s. \end{aligned}$$

210 for $c \in (0, \infty)$.

Proof. Consider item (iv) of Lemma 3 and assume $\mathcal{F}_{t-1} = \sigma(X_1, \dots, X_t)$, $Z_t = t^\delta$, $\delta > 1$,

$$U_t = \frac{\partial l_t(\theta_r)}{\partial \alpha_i} \quad \text{and} \quad W_n = \left[n E \left(\frac{\partial l_t(\theta_r)^2}{\partial \alpha_i} \right) \right]^{1/2},$$

where, by (9),

$$\frac{\partial l_t(\theta_r)}{\partial \alpha_i} = \frac{1}{2} \text{Tr} \left(X_t X_t' H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} H_t^{-1} - H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} \right).$$

To apply Theorem 3, we need to prove conditions L1-L4 below.

(L1) By the Chebyshev's inequality, we have

$$\begin{aligned} P(|U_t| > Z_t) &= P \left(\left| \frac{\partial l_t(\theta_r)}{\partial \alpha_i} \right| > t^\delta \right) \\ &\leq \frac{1}{t^{2\delta}} E \left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i}^2 \right). \end{aligned} \quad (\text{B.2})$$

Using item (iii) of Lemma 3,

$$\sum_{t=1}^{\infty} P(|U_t| > Z_t) \leq E \left(\frac{\partial l_1(\theta_r)}{\partial \alpha_i}^2 \right) \sum_{t=1}^{\infty} \frac{1}{t^{2\delta}} < \infty.$$

By the Borel-Cantelli Lemma,

$$P \left(\{ \omega : \mathbb{I}(|U_t| > t^\delta) = 1 \text{ i.o.} \} \right) = P \left(\{ \omega : |U_t| > t^\delta \text{ i.o.} \} \right) = 0$$

and thus

$$\lim_{n \rightarrow \infty} \frac{\sum_{t=1}^n U_t \mathbb{I}(|U_t| > Z_t) - E[U_t \mathbb{I}(|U_t| > Z_t) | \mathcal{F}_{t-1}]}{\sqrt{2W_n^2 \log \log W_n^2}} = 0 \quad a.s.$$

(L2)

$$\begin{aligned} E(U_t | \mathcal{F}_{t-1}) &= \frac{1}{2} \text{Tr} \left[E(X_t X_t' | \mathcal{F}_{t-1}) H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} H_t^{-1} - H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} \right] \\ &= \frac{1}{2} \text{Tr} \left[H_t H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} H_t^{-1} - H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} \right] \\ &= 0. \end{aligned}$$

Using item (iv) of Lemma 3,

$$E(E(U_t^2 | \mathcal{F}_{t-1})) = E(U_t^2) < \infty$$

and then, by the SLLN, we have

$$\lim_{n \rightarrow \infty} \frac{\sum_{t=1}^n E[U_t^2 | \mathcal{F}_{t-1}]}{n E \left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i}^2 \right)} = 1 \quad a.s.$$

By the Dominated Convergence Theorem,

$$\lim_{t \rightarrow \infty} E(U_t \mathbb{I}(|U_t| \leq n) | \mathcal{F}_{t-1}) = 0 \quad a.s.$$

Considering an arbitrary $\varepsilon > 0$, it is required to find a t -summable upper bound for

$$P \left[|E(U_t^2 | \mathcal{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | \mathcal{F}_{t-1})| > \varepsilon \right]$$

and apply the Borel-Cantelli Lemma to conclude

$$\lim_{t \rightarrow \infty} [E(U_t^2 \mathbb{I}(|U_t| \leq t^\delta) | \mathcal{F}_{t-1}) - E(U_t^2 | \mathcal{F}_{t-1})] = 0 \quad a.s.$$

and apply the Cesàro's Mean Theorem to conclude

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\sum_{t=1}^n E[U_t^2 \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}] - E[U_t \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}]^2}{W_n^2} &= \lim_{n \rightarrow \infty} \frac{\sum_{t=1}^n E[U_t^2 | \mathcal{F}_{t-1}]}{n E \left(\frac{\partial I_t(\theta_r)^2}{\partial \alpha_t} \right)} \\ &= 1 \quad a.s. \end{aligned}$$

215

By the generalized Chebyshev's inequality,

$$\begin{aligned} P \left[|E(U_t^2 | \mathcal{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | \mathcal{F}_{t-1})| > \varepsilon \right] \\ \leq \frac{1}{\varepsilon} E \left[|E(U_t^2 | \mathcal{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | \mathcal{F}_{t-1})| \right] \end{aligned}$$

and

$$\begin{aligned} E \left[|E(U_t^2 | \mathcal{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | \mathcal{F}_{t-1})| \right] \\ = E \left[|E(U_t^2 | \mathcal{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | \mathcal{F}_{t-1}) \pm E(U_t^2 I(|U_t| > t^\delta) | \mathcal{F}_{t-1})| \right] \\ = E \left[E(U_t^2 I(|U_t| > t^\delta) | \mathcal{F}_{t-1}) \right] \\ = E \left[U_t^2 I(|U_t| > t^\delta) \right] \\ \leq E \left[U_t^4 \right]^{1/2} E \left[I(|U_t| > t^\delta) \right]^{1/2} \\ = E \left[U_t^4 \right]^{1/2} P(|U_t| > t^\delta)^{1/2}. \end{aligned} \tag{B.3}$$

Using (B.2), (B.7) and (B.3), for suitable $c > 0$,

$$E \left[|E(U_t^2 | \mathcal{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | \mathcal{F}_{t-1})| \right] \leq c \frac{1}{t^\delta}$$

which is t -summable.

(L3) Using the notation $\dot{H}_t := \frac{\partial H_t}{\partial \alpha_i}$,

$$\begin{aligned}
0 &\leq E[U_t^4 \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}] \\
&\leq E[U_t^4 | \mathcal{F}_{t-1}] \\
&\leq E \left\{ \text{Tr} \left[X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} - H_t^{-1} \dot{H}_t \right]^4 | \mathcal{F}_{t-1} \right\} \\
&= E \left\{ \left[\text{Tr} \left(X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right) - \text{Tr} \left(H_t^{-1} \dot{H}_t \right) \right]^4 | \mathcal{F}_{t-1} \right\} \\
&= E \left\{ \text{Tr} \left(X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right)^4 \right. \\
&\quad - 4 \text{Tr} \left(X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right)^3 \text{Tr} \left(H_t^{-1} \dot{H}_t \right) \\
&\quad + 6 \text{Tr} \left(X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right)^2 \text{Tr} \left(H_t^{-1} \dot{H}_t \right)^2 \\
&\quad - 4 \text{Tr} \left(X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right) \text{Tr} \left(H_t^{-1} \dot{H}_t \right)^3 \\
&\quad \left. + \text{Tr} \left(H_t^{-1} \dot{H}_t \right)^4 | \mathcal{F}_{t-1} \right\}. \tag{B.4}
\end{aligned}$$

220

Also, by Lemma 3 (vi), for suitable $c \in (0, \infty)$,

$$\begin{aligned}
\left| \text{Tr} \left(X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right) \right| &= \left| \text{Tr} \left(H_t^{1/2} \varepsilon_t \left(H_t^{1/2} \varepsilon_t \right)' H_t^{-1} \dot{H}_t H_t^{-1} \right) \right| \\
&= \left| \text{Tr} \left(H_t^{1/2} \varepsilon_t \varepsilon_t' H_t^{1/2} H_t^{-1} \dot{H}_t H_t^{-1} \right) \right| \\
&= \left| \text{Tr} \left(\varepsilon_t \varepsilon_t' H_t^{-1/2} \dot{H}_t H_t^{-1/2} \right) \right| \\
&\leq \|\varepsilon_t \varepsilon_t'\| \left\| H_t^{-1/2} \right\|^2 \left\| \dot{H}_t \right\| \\
&\leq c \|\varepsilon_t \varepsilon_t'\| \left\| \dot{H}_t \right\| \tag{B.5}
\end{aligned}$$

and

$$\left| \text{Tr} \left(H_t^{-1/2} \dot{H}_t H_t^{-1/2} \right) \right| \leq c \left\| \dot{H}_t \right\|. \tag{B.6}$$

We used the relation $|\text{Tr}(ABC)| \leq \|A\| \|B\| \|C\|$. Thus,

$$\begin{aligned}
E[U_t^4 \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}] &\leq c^4 E \left(\|\varepsilon_t \varepsilon_t'\|^4 | \mathcal{F}_{t-1} \right) \left\| \dot{H}_t \right\|^4 \\
&\quad + 4c^4 E \left(\|\varepsilon_t \varepsilon_t'\|^3 | \mathcal{F}_{t-1} \right) \left\| \dot{H}_t \right\|^4 \\
&\quad + 6c^4 E \left(\|\varepsilon_t \varepsilon_t'\|^2 | \mathcal{F}_{t-1} \right) \left\| \dot{H}_t \right\|^4 \\
&\quad + 3c^4 \left\| \dot{H}_t \right\|^4 \\
&\leq c_1 E \left[\|\varepsilon_t \varepsilon_t'\|^2 + \|\varepsilon_t \varepsilon_t'\|^3 + \|\varepsilon_t \varepsilon_t'\|^4 + 1 | \mathcal{F}_{t-1} \right] \left\| \dot{H}_t \right\|^4. \tag{B.7}
\end{aligned}$$

Adapting the proof of Lemma A.2 of Comte & Lieberman [41] and using B5, we conclude that

$$E \sup_{\theta \in \Theta_k} \left[\left\| \frac{\partial H_t}{\partial \alpha_i}(\theta) \right\|^8 \right] < \infty. \quad (\text{B.8})$$

225

For $\delta_1 \in (\frac{1}{2}, 1)$ and suitable $c_2, c_4 \in (0, \infty)$, using the Chebyshev's and Jensen's inequalities, (B.7) and (B.8),

$$\begin{aligned} & \mathbb{P} \left[E[U_t^4 \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}] > t^{\delta_1} \right] \\ & \leq \mathbb{P} \left[E \left[\|\varepsilon_t \varepsilon'_t\|^2 + \|\varepsilon_t \varepsilon'_t\|^3 + \|\varepsilon_t \varepsilon'_t\|^4 + 1 | \mathcal{F}_{t-1} \right] \|\dot{H}_t\|^4 > \frac{t^{\delta_1}}{c_1} \right] \\ & \leq \frac{c_2^2}{t^{2\delta_1}} E \left\{ E \left[\left(\|\varepsilon_t \varepsilon'_t\|^2 + \|\varepsilon_t \varepsilon'_t\|^3 + \|\varepsilon_t \varepsilon'_t\|^4 + 1 \right) \|\dot{H}_t\|^4 | \mathcal{F}_{t-1} \right]^2 \right\} \\ & \leq \frac{c_2^2}{t^{2\delta_1}} E \left\{ E \left[\left(\|\varepsilon_t \varepsilon'_t\|^2 + \|\varepsilon_t \varepsilon'_t\|^3 + \|\varepsilon_t \varepsilon'_t\|^4 + 1 \right)^2 \|\dot{H}_t\|^8 | \mathcal{F}_{t-1} \right] \right\} \\ & \leq \frac{c_2^2}{t^{2\delta_1}} E \left\{ \left(\|\varepsilon_t \varepsilon'_t\|^2 + \|\varepsilon_t \varepsilon'_t\|^3 + \|\varepsilon_t \varepsilon'_t\|^4 + 1 \right)^2 \right\} E \left\{ \|\dot{H}_t\|^8 \right\} \\ & \leq \frac{c_3}{t^{2\delta_1}} E \left[\|\dot{H}_t\|^8 \right] \end{aligned}$$

and

$$\sum_{t=1}^{\infty} \mathbb{P} \left[E[U_t^4 \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}] > t^{\delta_1} \right] \leq \sum_{t=1}^{\infty} \frac{c_3}{t^{2\delta_1}} E \left[\|\dot{H}_t\|^8 \right] < \infty.$$

By the Borel-Cantelli Lemma,

$$\mathbb{P} \left[E[U_t^4 \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}] > t^{\delta_1} \quad i.o. \right] = 0$$

and then

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n \frac{E[U_t^4 \mathbb{I}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}]}{W_t^4} \leq \lim_{n \rightarrow \infty} \sum_{t=1}^n \frac{1}{t^{2-\delta_1}} < \infty \quad a.s.$$

(L4) $E \left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i} \right)^2 > 0$, otherwise $\frac{\partial l_t(\theta_r)}{\partial \alpha_i} \equiv 0$ a.s. and A_2 would be 0. Thus, using the stationarity,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{W_n}{W_{n+1}} &= \lim_{n \rightarrow \infty} \frac{\left[n E \left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i} \right)^2 \right]^{1/2}}{\left[(n+1) E \left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i} \right)^2 \right]^{1/2}} = 1 \quad a.s. \text{ and} \\ \lim_{n \rightarrow \infty} W_n &= E \left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i} \right)^2 \lim_{n \rightarrow \infty} \sqrt{n} = \infty. \end{aligned}$$

□

Lemma 6. Let $\mathbb{X} = \{X_t\}_{t \in \mathbb{N}}$ be a BEKK-GARCH(k), where conditions (8) and B1-B5 are true. Then,

$$E \sup_{\theta \in \Theta_k} \left[\left| \text{Tr}(\dot{H}_t(\theta)H_t^{-1}(\theta) - X_t X_t' H_t^{-1}(\theta) \dot{H}_t(\theta) H_t^{-1}(\theta)) \right| \right] < \infty,$$

230 where $\dot{H}_t := D_\theta H_t$.

Proof. Using Lemma 3 and (B.8), for a suitable $c \in (0, \infty)$,

$$\begin{aligned} E \sup_{\theta \in \Theta_k} \left[\left| \text{Tr}(\dot{H}_t H_t^{-1} - X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1}) \right| \right] &\leq E \sup_{\theta \in \Theta_k} \left[\left\| \dot{H}_t \right\| \left\| H_t^{-1} \right\| + \left\| X_t X_t' \right\| \left\| H_t^{-1} \right\|^2 \left\| \dot{H}_t \right\| \right] \\ &\leq E \sup_{\theta \in \Theta_k} \left[c \left\| \dot{H}_t \right\| + c^2 \left\| X_t X_t' \right\| \left\| \dot{H}_t \right\| \right] < \infty. \end{aligned}$$

□

Lemma 7. Let $\mathbb{X} = \{X_t\}_{t \in \mathbb{N}}$ be a BEKK-GARCH(r), of order r , $k \not\geq r$, θ_r its true parameter, $\hat{\theta}_k$ the MLE of θ_r . If conditions (8) and B1-B5 are true, then

$$\lim_{n \rightarrow \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,k}(\hat{\theta}_k)}{n} > 0 \text{ a.s.}$$

Proof. Assuming p such as $p \geq k$ and $p \geq r$,

$$\lim_{n \rightarrow \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,k}(\hat{\theta}_k)}{n} = \lim_{n \rightarrow \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,p}(\hat{\theta}_p) + \log L_{n,p}(\hat{\theta}_p) - \log L_{n,k}(\hat{\theta}_k)}{n}.$$

Applying Lemma 2 using the results above, we have

$$\lim_{n \rightarrow \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,p}(\hat{\theta}_p)}{n} = 0 \text{ a.s.}$$

By (9), we see that we just need to prove

$$\lim_{n \rightarrow \infty} \frac{\log L_{n,p}(\hat{\theta}_p) - \log L_{n,k}(\hat{\theta}_k)}{n} > 0 \text{ a.s.}$$

considering

$$\log L_{n,k}(\theta) = \sum_{t=1+k}^n l_t(\theta)$$

and

$$l_t(\theta) = \log \left(\frac{1}{\sqrt{(2\pi)^{m/2} \det(H_t)}} \exp \left(-\frac{1}{2} x_t' H_t^{-1} x_t \right) \right).$$

By Lemma 4, item (iii),

$$E [|l_t(\theta_r)|] < \infty$$

and then, using the SLLN,

$$\lim_{n \rightarrow \infty} \frac{\log L_{n,p}(\theta_r)}{n} = \lim_{n \rightarrow \infty} \frac{\sum_{t=1+\bar{k}}^n l_t(\theta_r)}{n} = E(l_1(\theta_r)) = c_1 < \infty \quad a.s.$$

235 By the Mean Value Theorem, for $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_p$, $s \in (0,1)$, sufficiently large n and $B_\delta(\theta_r)$ a sufficiently small neighborhood of θ_r ,

$$\begin{aligned} \left| \frac{\sum_{t=1+\bar{k}}^n l_t(\hat{\theta}_p)}{n} - \frac{\sum_{t=1+\bar{k}}^n l_t(\theta_r)}{n} \right| &= \left| \frac{\sum_{t=1+\bar{k}}^n D_{\dot{\theta}}^1 l_t(\dot{\theta})}{n} (\hat{\theta}_p - \theta_r) \right| \\ &\leq \sup_{\theta \in B_\delta(\theta_r)} \left\| \frac{\sum_{t=1+\bar{k}}^n D_{\dot{\theta}}^1 l_t(\dot{\theta})}{n} \right\| \|\hat{\theta}_p - \theta_r\| \end{aligned}$$

Using the SLLN, Lemma 6 and the strong consistency of $\hat{\theta}_p$,

$$\left| \frac{\sum_{t=1+\bar{k}}^n l_t(\hat{\theta}_p)}{n} - \frac{\sum_{t=1+\bar{k}}^n l_t(\theta_r)}{n} \right| \xrightarrow{a.s.} 0. \quad (\text{B.9})$$

And thus,

$$\lim_{n \rightarrow \infty} \frac{\log L_{n,p}(\hat{\theta}_p)}{n} = \lim_{n \rightarrow \infty} \frac{\log L_{n,p}(\theta_r)}{n} = c_1. \quad (\text{B.10})$$

Besides that, $\Theta_k \subset \Theta_p$ and $\hat{\theta}_k$ is the MLE of θ_r , thus

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\log L_{n,k}(\hat{\theta}_k)}{n} &\leq \lim_{n \rightarrow \infty} \frac{\log L_{n,p}(\hat{\theta}_p)}{n} \\ &= c_1 \quad a.s. \end{aligned}$$

240 and then,

$$\lim_{n \rightarrow \infty} \frac{\log L_{n,k}(\hat{\theta}_k)}{n} \leq \limsup_{n \rightarrow \infty} \frac{\sum_{t=1+\bar{k}}^n l_t(\hat{\theta}_k)}{n} = c_2 \leq c_1 \quad a.s.$$

Let n_i be a subsequence of n such as

$$\lim_{n \rightarrow \infty} \frac{\sum_{t=1+\bar{k}}^{n_i} l_t(\hat{\theta}_k)}{n_i} = c_2 \quad a.s.$$

Using that Θ_k is compact, assume n_j a subsequence of n_i such as

$$\hat{\theta}_k(n_j) \rightarrow \bar{\theta}_k \in \Theta_k \quad a.s.$$

And then,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\log L_{n,k}(\hat{\theta}_k)}{n} &\leq \limsup_{n \rightarrow \infty} \frac{\log L_{n,k}(\hat{\theta}_k)}{n} \\ &= \lim_{n_j \rightarrow \infty} \frac{\log L_{n_j}(\hat{\theta}_k(n_j))}{n_j} \quad a.s. \end{aligned}$$

Applying the same argument used in (B.9), we have

$$\lim_{n_i \rightarrow \infty} \frac{\log L_{n_j}(\bar{\theta}_k)}{n_j} = E(l_1(\bar{\theta})) \quad and$$

245

$$\left| \frac{\sum_{t=1+k}^n l_t(\hat{\theta}_k)}{n} - \frac{\sum_{t=1+k}^n l_t(\bar{\theta})}{n} \right| \xrightarrow{a.s.} 0.$$

Additionally,

$$\lim_{n \rightarrow \infty} \left[\frac{\log L_{n,p}(\hat{\theta}_p)}{n} - \frac{\log L_{n,k}(\hat{\theta}_k)}{n} \right] \geq E \left[\log \left(\frac{f(\bar{\theta}_k)}{f(\theta_r)} \right) \right].$$

By other hand,

$$E \left[\log \left(\frac{f(\bar{\theta}_k)}{f(\theta_r)} \right) \right]$$

is the Kullback-Leibler divergence, which is positive if $f(\bar{\theta}_k) \neq f(\theta_r)$, and, as $\theta_r \notin \Theta_k \subseteq \mathbb{R}^{\gamma(k)}$, we have $\theta_r \neq \bar{\theta}_k$ and then, by B3, $f(\bar{\theta}_k) \neq f(\theta_r)$. Thus, we conclude

$$\lim_{n \rightarrow \infty} \left[\frac{\log L_{n,k}(\hat{\theta}_r)}{n} - \frac{\log L_{n,k}(\hat{\theta}_k)}{n} \right] \geq E \left[\log \left(\frac{f(\bar{\theta}_k)}{f(\theta_r)} \right) \right] = c > 0.$$

250

□

Lemmas 3 and 4 provides A1-A3 and Lemmas 5 and 7 provides, respectively, A4 and A5. The EDC estimator's consistency is established using Theorem 1.

References

- [1] M. S. Bartlett, The frequency goodness of fit test for probability chains,
255 Proceedings of the Cambridge Philosophical Society 47 (1951) 86–95.
- [2] P. G. Hoel, A test for Markoff chains, *Biometrika* 41 (3/4) (1954) 430–433.
- [3] I. J. Good, The likelihood ratio test for Markoff chains, *Biometrika* 42 (3/4)
(1955) 531–533.
- [4] P. Billingsley, Statistical methods in Markov chains, *The Annals of Math-*
260 *ematical Statistics* 32 (1) (1961) 12–40.
- [5] T. W. Anderson, L. A. Goodman, Statistical inference about Markov
chains, *The Annals of Mathematical Statistics* 28 (1) (1957) 89–110.
- [6] M. H. Quenouille, A large-sample test for the goodness of fit of autore-
gressive schemes, *Journal of the Royal Statistical Society* 110 (2) (1947)
265 123–129.
- [7] P. Whittle, *Hypothesis Testing in Time-Series Analysis*, Almqvist and Wik-
sells, Uppsala, 1951.
- [8] P. Whittle, Some recent contributions to the theory of stationary processes:
A Study in the analysis of stationary time series., Almqvist and Wiksells,
270 Uppsala, 1954.
- [9] M. S. Bartlett, D. V. Rajalakshman, Goodness of fit tests for simultaneous
autoregressive series, *Journal of the Royal Statistical Society* 15 (1) (1953)
107–124.
- [10] H. Akaike, A new look at the statistical model identification, *Automatic*
275 *Control, IEEE Transactions on* 19 (6) (1974) 716–723.
- [11] G. Schwarz, Estimating the dimension of a model, *The Annals of Statistics*
6 (2) (1978) 461–464.

- [12] H. Akaike, A bayesian extension of the minimum AIC procedure of autoregressive model fitting, *Biometrika* 66 (1979) 237–242.
- 280 [13] E. J. Hannan, B. G. Quinn, The determination of the order of an autoregression, *Journal of the Royal Statistical Society* 41 (2) (1979) 190–195.
- [14] E. J. Hannan, The estimation of the order of an arma process, *The Annals of Statistics* 8 (1980) 1071–1081.
- [15] T. Ozaki, On the order determination of ARIMA models, *Journal of Applied Statistics* 26 (1977) 290–301.
- 285 [16] Y. Ogata, Maximum likelihood estimates of incorrect Markov models for time series and the derivation of AIC, *Journal of Applied Probability* 17 (1980) 59–72.
- [17] B. Choi, *ARMA Model Identification*, Springer, New York, 1992.
- 290 [18] H. Tong, Determination of the order of a Markov chain by Akaike’s information criterion, *Journal of Applied Probability* 12 (3) (1975) 488–497.
- [19] R. W. Katz, On some criteria for estimating the order of a Markov chain, *Technometrics* 23 (3) (1981) 243–249.
- [20] I. Csiszár, P. C. Shields, The consistency of the BIC Markov order estimator, *The Annals of Statistics* 28 (6) (2000) 1601–1619.
- 295 [21] A. M. Polansky, Detecting change-points in Markov chains, *Computational Statistics & Data Analysis* 51 (2007) 6013–6026.
- [22] L. Finesso, Consistent estimation of the order for Markov and hidden Markov chains, Ph.D. thesis, University of Maryland (1990).
- 300 [23] A. W. Hughes, M. L. King, K. T. Kwek, Selecting the order of an ARCH model, *Economics Letters* 83 (2) (2004) 269–275.
- [24] A. E. Raftery, A model for high-order Markov chains, *Journal of the Royal Statistical Society* 47 (3) (1985) 528–539.

- [25] L. Zhao, C. Dorea, C. Gonçalves, On determination of the order of a Markov chain, *Statistical Inference for Stochastic Processes* 4 (3) (2001) 273–282.
- [26] C. C. Y. Dorea, Optimal penalty term for EDC Markov chain order estimator, *Annales de l’Institut de Statistique de l’Universite de Paris* 52 (3) (2008) 15–26.
- [27] P. A. A. Resende, Análise comparativa de estimadores da ordem de cadeias de Markov, Master’s thesis, Universidade de Brasília (2009).
- [28] R. Nishii, Maximum likelihood principle and model selection when the true model is unspecified, *Journal of Multivariate Analysis* 27 (2) (1988) 392–403.
- [29] I. V. Basawa, C. C. Heyde, Asymptotic properties of maximum likelihood estimators for stochastic processes, *The Indian Journal of Statistics* 38 (1976) 259–270.
- [30] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica* 50 (4) (1982) 987–1007.
- [31] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31 (3) (1986) 307–327.
- [32] R. F. Engle, V. K. Ng, Measuring and testing the impact of news on volatility, *The Journal of Finance* 48 (5) (1993) 1749–1778.
- [33] D. B. Nelson, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59 (2) (1991) 349–370.
- [34] R. F. Engle, K. F. Kroner, Multivariate simultaneous generalized ARCH, *Econometric Theory* 11 (1) (1995) 122–150.
- [35] T. Bollerslev, R. F. Engle, J. M. Wooldridge, A Capital Asset Pricing Model with Time-Varying Covariances, *Journal of Political Economy* 96 (1) (1988) 116–131.

- [36] T. Bollerslev, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *Review of Economics and Statistics* 72 (3) (1990) 498–505.
- [37] R. F. Engle, V. K. Ng, M. Rothschild, Asset pricing with a Factor-ARCH
335 covariance structure: Empirical estimates for Treasury bills, *Journal of Econometrics* 45 (1990) 213–237.
- [38] F. Boussama, F. Fuchs, R. Stelzer, Stationarity and geometric ergodicity of BEKK multivariate GARCH models, *Stochastic Processes and their Applications* 121 (10) (2011) 2331 – 2360.
- [39] R. Stelzer, On the relation between the VEC and BEKK multivariate
340 GARCH models, *Econometric Theory* 24 (04) (2008) 1131–1136.
- [40] F. Boussama, Ergodicité, mélange et estimation dans les modèles garch, Ph.D. thesis, Université 7 Paris (1998).
- [41] F. Comte, O. Lieberman, Asymptotic theory for multivariate GARCH pro-
345 cesses, *Journal of Multivariate Analysis* 84 (2003) 61–84.
- [42] T. Jeantheau, Strong consistency of estimators for multivariate ARCH models, *Econometric Theory* 14 (1) (1998) 70–86.
- [43] S. P. Meyn, R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, London, 1993.
- [44] P. Hall, C. C. Heyde, *Martingale limit theory and its application*, Academic
350 Press, New York, 1980.
- [45] C. M. Hafner, A. Preminger, On asymptotic theory for multivariate GARCH models, *Journal of Multivariate Analysis* 100 (2009) 2044–2054.
- [46] C. M. Hafner, A. Preminger, Asymptotic theory for a FACTOR GARCH
355 model, *Econometric Theory* 25 (2009) 336–363.